Matrix product representation and synthesis for random vectors: Insight from statistical physics

Florian Angeletti*, Eric Bertin, Patrice Abry,

Université de Lyon, Laboratoire de Physique, Ecole Normale Supérieure de Lyon, CNRS
46 allée d'Italie, F-69007 Lyon, France
ens-lyon.fr/PHYSIQUE/, firstname.lastname@ens-lyon.fr

Abstract—Inspired from modern out-of-equilibrium statistical physics models, a matrix product based framework permits the formal definition of random vectors (and random time series) whose desired joint distributions are a priori prescribed. Its key feature consists of preserving the writing of the joint distribution as the *simple* product structure it has under independence, while inputing controlled dependencies amongst components: This is obtained by replacing the product of distributions by a product of matrices of distributions. The statistical properties stemming from this construction are studied theoretically: The landscape of the attainable dependence structure is thoroughly depicted and a stationarity condition for time series is notably obtained. The remapping of this framework onto that of Hidden Markov Models enables us to devise an efficient and accurate practical synthesis procedure. A design procedure is also described permitting the tuning of model parameters to attain targeted properties. Pedagogical well-chosen examples of times series and multivariate vectors aim at illustrating the power and versatility of the proposed approach and at showing how targeted statistical properties can be actually prescribed.

Index Terms—Random vectors, Time Series, Joint Distribution, A priori Prescription, Numerical Simulation, Matrix Product, Hidden Markov Model, Statistical Physics

EDICS Category: SSP-SNMD SSP-NGAU SSP-NSSP

I. INTRODUCTION

In modern signal processing, it is very often needed that synthetic data are produced numerically in fast and efficient manners with (some of) their statistical properties being a priori prescribed as closely as desired from selected targets (marginal distributions, covariance, spectrum, multivariate distribution,...). This is for instance the case when the performance of newly developed statistical analysis need to be assessed. The theoretical derivation of such performance may turn too difficult to achieve, specially when it is intended to apply such tools to real-word data, whose properties are not well known. Instead, one can resort to Monte Carlo simulations: Performance are derived from averages over independent realizations of synthetic data, that are designed to resemble as closely as possible the real-world data. Another example stems from Bayesian estimation procedures, that generally involve Monte Carlo Markov Chain [1] or variational based resolution schemes (cf. e.g., [2], [3]) to generate numerically independent copies of hyper-parameters drawn from (possibly complicated) distributions, derived from the Bayesian formalism.

For the numerical synthesis of Gaussian time series, with prescribed covariance function, the so-called *Circulant Embeded Matrix* synthesis procedure [4]–[6] is nowadays considered as the state-of-the-art solution and is currently widely used. It has then been extended to the synthesis of multivariate Gaussian time series, with prescribed auto- and cross-covariance, notably in [7], [8] (see also [9] for variations). The synthesis of non-Gaussian time series turns far more complicated as the prescription of the full joint distributions turns very difficult to achieve while that of the sole marginal distributions and covariance functions does not uniquely define the process. Several approaches were proposed to address such issues (cf. e.g., [10], [11] and references therein for reviews). Other strategies are based on surrogate techniques: Starting from the original real-world data of interest, surrogate copies are obtained by randomizing one attribute of the data (e.g., the phase of its Fourier transform) while maintaining another fixed (e.g., the amplitude of the Fourier transform) (cf. e.g., [12]). Recently, an *optimal transport* procedure was proposed aiming at iteratively modifying the joint distribution of random vectors or time series to attain a given target [13]. The general framework of Markov Chain simulation offers an alternative and broad class of solutions, focusing on the modelling of local dynamical properties, while not explicitly putting the emphasis on a direct prescription of the joint distributions of the process. Markov Chain schemes are also widely used to synthesize independent realizations of random vectors with prescribed properties (cf. e.g., [14] for a review).

The present contribution takes place in this long tradition of synthetic data design in signal processing, but is however rooted in a very different scientific field, that of statistical physics. Indeed, inspired from the exact solutions of stochastic out-of-equilibrium models describing particle diffusion on a one-dimensional lattice with volume exclusion, like the Asymmetric Simple Exclusion Processes (ASEP) [15]–[17], the design procedure proposed here is founded on the request that the joint distribution of the random vector or random times series $\underline{X}_N = x_1, x_2, \ldots, x_N$ must be written as a product $P_{\underline{X}} \propto \prod_{k=1}^{N} R_d(x_k)$, as in the case of independent components. However, the R_ds are no longer univariate distributions, but instead d dimensional matrices of valid unnormalized distributions. In statistical physics, this matrix product form of the joint distribution was envisaged as a practical ansatz to find exact solutions of the master equation associated e.g.

to the ASEP model. In signal processing, this enables us to devise a theoretically powerful and practically versatile and efficient procedure to define and generate numerically random vectors or random times series, with a priori prescribed statistical properties. A preliminary and partial attempt relying on specific choices of matrices and focusing on stationary time series only was presented in [18]. It is here extended to a more general and versatile setting.

Definitions of this matrix product joint distribution framework and the general consequences in terms of statistical properties are derived in Section II. Notably, a general condition to ensure stationarity of the time series is obtained. In Section III, the dependence structures that can be achieved from this model are studied in-depth and the practical tuning of the time-scales that can be input in these dependence structures is established in details. In Section IV, it is first shown how this matrix product framework can be explicitly recast into that of Hidden Markov Models, hence providing us with a fast and efficient practical synthesis procedure; and it is, second, explained how targeted marginal distributions and other statistical properties can be attained. Section V consists of two different sets of pedagogical examples aiming at illustrating the efficiency and versatility of the tool: First, sets of time series whose marginal distributions, covariance functions and some higher order statistics are jointly prescribed, are produced; Second, sets of random vectors, with different marginal distributions are synthesized and it is shown how to tune the correlation and fourth order statistics amongst the components of such vectors while maintaining fixed the marginal distributions.

II. JOINT PROBABILITY AS MATRIX PRODUCT

A. Definitions and general setting

Inspired from out-of-equilibrium statistical physics models (cf. e.g. [15]–[17]), a general framework for the design of joint distribution functions for random vectors $\underline{X}_N$, of size N is devised. It is based on product of matrices and relies on 4 different key ingredients defined below.

First, let $d \in \mathbb{N}^*$. Second, let $\mathcal{A}$ denote a fixed non-zero non-random $d \times d$ matrix, with positive entries, and $\mathcal{L}(M)$ an associated linear form defined (for any matrix M) as:

$$\mathcal{L}(M) = \mathsf{tr}\left(\mathcal{A}^T M\right). \tag{1}$$

Third, let $\mathcal{E}$ denote a fixed non-zero non-random $d \times d$ matrix, with positive entries $\mathcal{E}_{i,j}$. For reasons made clear in Section III, $\mathcal{E}$ will be here referred to as the *structure matrix*. Fourth, let Matrix $\mathcal{P}$ consists of a $d \times d$ set of valid (normalized) distribution functions $\{\mathcal{P}_{i,j}(x)\}_{i=1,\dots,d;j=1,\dots,d}$.

$$R_d(x) = \mathcal{E} \otimes \mathcal{P}(x), \tag{2}$$

where $\otimes$ denotes entry-wise matrix multiplication. Let $\underline{X} \equiv \{X_n\}_{1 \leq n \leq N}$ denote the random vector explicitly defined via its joint distribution:

$$\mathbb{P}_{\underline{X}}(x_1, \dots, x_N) = \frac{\mathcal{L}\left(\prod_{k=1}^N R_d(x_k)\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}, \tag{3}$$

where $\prod_{k=1}^N R_d(x_k) = R_d(x_1) \dots R_d(x_N)$ denotes the oriented product (i.e., the order of the factors is fixed). The joint

choice of positive entries for matrices $\mathcal{A}$ and $\mathcal{E}$, and of valid distribution functions in $\mathcal{P}$ is sufficient to ensure that Eq. (3) defines a valid joint distribution function.

B. Marginal distributions, moments and dependence

From these definitions, using commutativity of integration and matrix product, a number of statistical properties of $\underline{X}_N$ can be analytically derived.

Univariate or marginal distributions take explicit forms:

$$\begin{aligned}
\mathbb{P}_k(X_k = x_k) &= \frac{\int \mathcal{L}\left(\prod_{i=1}^N R_d(x_i)\right) \prod_{i \neq k} dx_i}{\mathcal{L}\left(\mathcal{E}^N\right)} \\
&= \frac{\mathcal{L}\left(\int R_d(x_1)dx_1 \dots R_d(x_k) \dots R_d(x_N)dx_N\right)}{\mathcal{L}\left(\mathcal{E}^N\right)} \\
&= \frac{\mathcal{L}\left(\mathcal{E}^{k-1} R_d(x_k) \mathcal{E}^{N-k}\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}.
\end{aligned} \tag{4}$$

This shows that the marginal distributions necessarily consist of weighted linear combinations of the $\mathcal{P}_{i,j}$,

$$\mathbb{P}_k(X_k = x) = \sum_{i,j} c_{i,j,k} \mathcal{P}_{i,j}(x), \tag{5}$$

where the $c_{i,j,k}$ are functions of $\mathcal{A}$ and $\mathcal{E}$, satisfying, for all k, $\sum_{i,j} c_{i,j,k} = 1$.

Further, let us define the collection of matrices, $M(q) = \int_{\mathbb{R}} x^q R_d(x)dx = \mathcal{E} \otimes \int_{\mathbb{R}} x^q \mathcal{P}(x)dx$, whose entries read:

$$M(q)_{i,j} = \mathcal{E}_{i,j} \int_{\mathbb{R}} x^q \mathcal{P}_{i,j}(x)dx. \tag{6}$$

Univariate moments also take closed-form expressions:

$$\mathbb{E}\left[X_k^q\right] = \frac{\mathcal{L}(\mathcal{E}^{k-1} M(q) \, \mathcal{E}^{N-k})}{\mathcal{L}(\mathcal{E}^N)}. \tag{7}$$

The p-variate distributions (with $k_1 < \dots < k_p$) can also be made explicit,

$$\begin{aligned}
&\mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) = \\
&\frac{\mathcal{L}\left(\mathcal{E}^{k_1 - 1}\left(\prod_{r=1}^{p-1} R_d(x_{k_r}) \mathcal{E}^{k_{r+1} - k_r - 1}\right) R_d(x_{k_p}) \mathcal{E}^{N-k_p}\right)}{\mathcal{L}\left(\mathcal{E}^N\right)},
\end{aligned} \tag{8}$$

as well as the p-variate moments (with q_r the order associated to entry k_r):

$$\begin{aligned}
&\mathbb{E}\left[\prod_{r=1}^p X_{k_r}^{q_r}\right] = \\
&\frac{\mathcal{L}\left(\mathcal{E}^{k_1 - 1}\left(\prod_{r=1}^{p-1} M(q_r) \mathcal{E}^{k_{r+1} - k_r - 1}\right) M(q_p) \mathcal{E}^{N-k_p}\right)}{\mathcal{L}(\mathcal{E}^N)}.
\end{aligned} \tag{9}$$

Eq. (9) constitutes a key result with respect to applications as it clearly shows that the joint statistics of order q of $\underline{X}_N$ can be prescribed by the sole selection of matrices $M(q)$. This will be explicitly used in Section V.

C. Stationarity condition

Eq. (3) shows that the in general non-stationary nature of $\underline{X}_N$ stems from the non-commutativity of the matrix product. To enforce stationarity in $\underline{X}_N$, a commutativity property for $\mathcal{A}$ and $\mathcal{E}$ must be added,

$$[\mathcal{A}^T, \mathcal{E}] \equiv \mathcal{A}^T \mathcal{E} - \mathcal{E}\mathcal{A}^T = 0, \qquad (10)$$

which ensures that $\forall x, \mathcal{L}\left(\mathcal{E}R_d(x)\right) = \mathcal{L}\left(R_d(x)\mathcal{E}\right)$.

Under Eq. (10), the marginal distribution (cf. Eq. (5)) simplify to:

$$\mathbb{P}_S(X = x) = \frac{\mathcal{L}\left(R_d(x)\mathcal{E}^{N-1}\right)}{\mathcal{L}\left(\mathcal{E}^N\right)} = \sum_{i,j} c_{i,j} \mathcal{P}_{i,j}(x), \quad (11)$$

and p-variate distributions and moments become (cf. Eq. (8) and Eq. (9))

$$\mathbb{P}(X_{k_1} = x_{k_1}, \ldots, X_{k_p} = x_{k_p}) =$$
$$\frac{\mathcal{L}\left(\left(\prod_{r=1}^{p-1} R_d(x_{k_r})\mathcal{E}^{k_{r+1}-k_r-1}\right) R_d(x_{k_p})\mathcal{E}^{N-(k_p-k_1)-1}\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}, \qquad (12)$$

$$\mathbb{E}\left[\prod_{r=1}^p X_{k_r}^{q_r}\right] =$$
$$\frac{\mathcal{L}\left(\left(\prod_{r=1}^{p-1} M(q_r)\mathcal{E}^{k_{r+1}-k_r-1}\right) M(q_p)\mathcal{E}^{N-(k_p-k_1)-1}\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}. \qquad (13)$$

These relations clearly indicate that the vector $\{X_n\}_{1 \le n \le N}$ can now be regarded as a stationary time series: All joint statistics depend only on time differences, $k_{r+1} - k_r$.

The sole matrix $\mathcal{A}$ satisfying Eq. (10) for all matrices $\mathcal{E}$, is the identity matrix, in which case $\mathcal{L}$ consists of the trace operator. However, the trace operator also induces automatically a circular correlation structure, $l \ge 2k$, $\mathbb{E}X_k X_l = \mathbb{E}X_k X_{N+2k-l}$, a highly undesirable consequence for application purposes.

Alternatively, stationarity can be obtained by choosing jointly specific pairs $(\mathcal{A}, \mathcal{E})$, such as

$$\mathcal{A}_{i,j} = (1/d) \qquad (14)$$

and $\mathcal{E}$ a so-called *doubly stochastic* matrix, defined as:

$$\forall i, j, \quad \sum_k \mathcal{E}_{k,j} = \sum_k \mathcal{E}_{i,k} \equiv 1. \qquad (15)$$

Indeed, such choices yield

$$\forall k, M, \quad \mathcal{L}\left(M\mathcal{E}^k\right) = \mathcal{L}\left(M\right),$$

that leads to further simplifications of Eq. (11) to Eq. (13). The marginal and partial distributions of $\underline{X}_N$ no longer explicitly depend on the sample size N, also the marginal distribution is independent of $\mathcal{E}$:

$$\mathbb{P}_k(X_k = x_k) = \mathcal{L}\left(R_d(x_k)\right), \qquad (16)$$

$$\mathbb{P}(X_{k_1} = x_{k_1}, \ldots, X_{k_p} = x_{k_p}) =$$
$$\mathcal{L}\left(\left(\prod_{r=1}^{p-1} R_d(x_{k_r})\mathcal{E}^{k_{r+1}-k_r-1}\right) R_d(x_{k_p})\right), \qquad (17)$$

$$\mathbb{E}\left[\prod_{r=1}^p X_{k_r}^{q_r}\right] = \mathcal{L}\left(\left(\prod_{r=1}^{p-1} M(q_r)\mathcal{E}^{k_{r+1}-k_r-1}\right) M(q_p)\right). \qquad (18)$$

Elaborating on a preliminary work (cf. [18]), this particular setting will be used in Section V-A to design efficiently stationary time series.

III. DEPENDENCE STRUCTURE

Eqs. (8) and (9) indicate that Matrix $\mathcal{E}$ essentially controls the dependence structure within Vector $\underline{X}_N$, hence its name, via the collection of its powers $\mathcal{E}^n$, $n = 1, \ldots, N$, while the matrices $M(q)$ fix the amplitudes of the dependencies at order q. Analyzing the forms possibly taken by the $\mathcal{E}^n$ is hence crucial to understand the potential dependence structures of $\underline{X}_N$ achievable in this framework. Notably, the classical distinction between diagonalizable and non-diagonalizable $\mathcal{E}$ plays a crucial role. This is studied in detail in this section for the correlation structure.

A. Diagonalizable structure matrix

Let us assume that $\mathcal{E}$ is diagonalizable, with eigenvalues $\lambda_1, \ldots, \lambda_r$. Then, $\mathcal{E}$ can be decomposed into r sub-matrices $E_1, \ldots, E_r$ such that

$$\mathcal{E}^k = \sum_{i=1}^r \lambda_i^k E_i, \qquad (19)$$

and Eq. (4) can be rewritten as:

$$\mathbb{P}_k(X_k = x) = \frac{\sum_{i,j} \lambda_i^{k-1} \lambda_j^{N-k} \mathcal{L}\left(E_i R_d(x) E_j\right)}{\sum_{i=1}^r \lambda_i^N \mathcal{L}\left(E_i\right)}, \qquad (20)$$

which explicits the dependence in k and is reminiscent of Eq. (5) given that the $\mathcal{L}\left(E_i R_d(x) E_j\right)$ consist of linear combinations of the $\mathcal{P}_{l,m}(x)$. Further, for the 2-sample statistics, or covariance function, Eq. (9) simplifies to:

$$\mathbb{E}\left[X_k X_l\right] = \frac{\sum_{i,j,m} \lambda_i^{k-1} \lambda_j^{N-l} \lambda_m^{k-l-1} \mathcal{L}\left(E_i M(1) E_m M(1) E_j\right)}{\sum_{i=1}^r \lambda_i^N \mathcal{L}\left(E_i\right)}. \qquad (21)$$

Assuming stationarity, i.e., Eq. (10), the relation above further reduces to:

$$\mathbb{E}\left[X_k X_l\right] = \frac{\sum_{j,m} \lambda_j^{N-2} \left(\frac{\lambda_m}{\lambda_j}\right)^{k-l-1} \mathcal{L}\left(M(1) E_m M(1) E_j\right)}{\sum_{i=1}^r \lambda_i^N \mathcal{L}\left(E_i\right)}, \qquad (22)$$

which shows that the covariance function consists of the sum of weighted exponential functions $\exp(-(k - l - 1)(\ln \lambda_j - \ln \lambda_m))$, with at most $r_M(r_M - 1)/2 = \binom{r_M}{2}$ characteristic time scales, $\tau_{j,m} = (\ln |\lambda_j| - \ln |\lambda_m|)^{-1}$, where r_M stands for the number of eigenvalues of Matrix $\mathcal{E}$ with different modulus. A preliminary study of the covariance function in the stationary and diagonalizable case has been devised in [18] and an example is worked out in Section V-A1.

Without assuming stationarity, this exponential decrease of the covariance function still holds. Indeed, let us assume that $\mathcal{L}\left(E_1\right) \ne 0$ and that the norm of λ_1 is strictly larger than the norm of the other eigenvalues. Then, the normalisation term $\mathcal{L}\left(\mathcal{E}^N\right)$ can be approximated in the limit $N \to +\infty$ as

$\mathcal{L}\left(\mathcal{E}^N\right) \sim \lambda_1^N \mathcal{L}\left(E_1\right)$. Combining this asymptotic form with Eq. (4) yields:

$$\mathbb{P}_k(X_k = x) = \sum_{i,j} \lambda_i^{k-1} \lambda_j^{N-k} \frac{\mathcal{L}\left(E_i R_d(x) E_j\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}$$
$$\sim \frac{\mathcal{L}\left(E_1 R_d(x) E_1\right)}{\lambda_1 \mathcal{L}\left(E_1\right)}, \quad (23)$$

and combining this result with Eq. (21) leads to:

$$\mathbb{E}\left[X_k X_l\right] \sim \sum_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k-l-1} \frac{\mathcal{L}\left(E_1 M(1) E_i M(1) E_1\right)}{\mathcal{L}\left(E_1\right) \lambda_1^2} \quad (24)$$

These equations show that using a diagonalizable $\mathcal{E}$, with a dominant eigenvalue, implies that asymptotically, i.e., in the limit $N \to +\infty$, each component of Vector $\underline{X}_N$ shares the same univariate distribution, and that the autocovariance functions reads as a sum of exponential functions, depending only on $|k - l|$. In the limit $N \to +\infty$, Vector $\underline{X}_N$ is hence asymptotically quasi-stationary.

B. Non-diagonalizable structure matrices

Non-diagonalizable matrices $\mathcal{E}$ draw a very different landscape. For illustration, let us consider the case where $\mathcal{E} = I_d + H$ with I_d is the Identity matrix and H any nilpotent matrix of order $p + 1$ (i.e., $H^{p+1} \equiv 0$ while $H^k \neq 0$, when $1 \le k \le p$), chosen as a standard example of non diagonalizable matrix. Then, for any $k \ge p$, one has:

$$\mathcal{E}^k = \sum_{j=0}^{p} \binom{k}{j} H^j,$$

which combined with Eq. (4) yields

$$\mathbb{P}_k(X_k = x_k) = \frac{\sum_{i,j \le p} \binom{k-1}{i}\binom{N-k}{j} \mathcal{L}\left(H^i R_d(x) H^j\right)}{\sum_{i=1}^{p} \binom{N}{i} \mathcal{L}\left(H^i\right)} \quad (25)$$

and, for the covariance function:

$$\mathbb{E}\left[X_k X_l\right] =$$
$$\frac{\sum_{i,j,m \le p} \binom{k-1}{i}\binom{k-l-1}{m}\binom{N-l}{j} \mathcal{L}\left(H^i M(1) H^m M(1) H^j\right)}{\sum_{i=1}^{p} \binom{N}{i} \mathcal{L}\left(H^i\right)} \quad (26)$$

To gain a better grasp of what these equations imply, let us study their asymptotic behaviors. Using

$$\binom{N}{p} \sim \frac{N^p}{p!}, N \to +\infty$$

and the assumption that $\mathcal{L}\left(H^p\right) \neq 0$ lead to

$$\mathcal{L}\left(\mathcal{E}^N\right) \sim \frac{N^p}{p!} \mathcal{L}\left(H^p\right)$$

and, with the assumption that k diverges with N, to

$$\mathbb{P}_k(X_k = x) = \sum_{i,j \le p} \binom{k-1}{i}\binom{N-k}{j} \frac{\mathcal{L}\left(H^i R_d(x) H^j\right)}{\mathcal{L}\left(\mathcal{E}^N\right)}$$
$$\sim \sum_{i+j=p} \binom{p}{i}\left(\frac{k}{N}\right)^i \left(1 - \frac{k}{N}\right)^j \frac{\mathcal{L}\left(H^j R_d(x) H^j\right)}{\mathcal{L}\left(H^p\right)}. \quad (27)$$

Compared to Eq. (20), this indicates that the marginal distribution of component k reads as a mixture of distributions, with weights depending on the relative position k/N, rather than the absolute position k, as is the case for diagonalizable $\mathcal{E}$.

The covariance function also turns quite different from the diagonalizable case,

$$\mathbb{E}\left[X_k X_l\right] \approx \sum_{i+j+m=p}$$
$$\left[\frac{p!}{i!m!j!}\left(\frac{k}{N}\right)^i \left(\frac{k-l}{N}\right)^j \left(1 - \frac{l}{N}\right)^m \frac{\mathcal{L}\left(H^i M_1 H^j M_1 H^m\right)}{\mathcal{L}\left(H^p\right)}\right] \quad (28)$$

with the occurrence of algebraic terms, $\left(\frac{k-l}{N}\right)^j$, that indicate long-range correlations developing along the whole vector $\underline{X}_N$.

This specific choice for non diagonalizable $\mathcal{E}$ enables us to figure out that, combining a block diagonal Jordan reduction of $\mathcal{E}$ and results obtained in the diagonalizable case, the covariance function consists, in the very general case, of a weighted sum of algebraic and exponential decreases.

IV. DESIGN AND SYNTHESIS

The elegant and compact definition of $\underline{X}_N$ via its joint distribution $\mathbb{P}(\underline{X}_N)$ in Eq. (3) gives little hints on how to construct random vectors with desired prescribed properties or on how to synthesize them numerically. The present section addresses such issues: First, given that $d, \mathcal{A}, \mathcal{E}$ and $\mathcal{P}$ are chosen, Eq. (3) is reformulated into the framework of Hidden Markov Models and the corresponding numerical synthesis algorithm is devised; Second, given $d, \mathcal{A}, \mathcal{E}$, an algorithm for constructing a Matrix $\mathcal{P}$ with prescribed marginal distributions and dependence structures is detailed.

A. Hidden Markov Chain

First, generalizing the fact that the entries of the matrix product (ABC) reads $(ABC)_{i,j} = \sum_{k,l} a_{i,k} b_{k,l} c_{l,j}$, enables us to recast Eq. (3) into:

$$\mathbb{P}(x_1, \ldots, x_n) = \sum_{\underline{\Gamma}} \kappa(\underline{\Gamma}) \prod_{k=1}^{N} \mathcal{P}(x_k)_{\Gamma_{k-1}, \Gamma_k}, \quad (29)$$

with $\underline{\Gamma} \equiv \{\Gamma_0, \ldots, \Gamma_k, \ldots, \Gamma_N\} \in [1, \ldots, d]^{N+1}$ and

$$\kappa(\underline{\Gamma}) = \frac{\mathcal{A}_{\Gamma_0, \Gamma_N}}{\mathcal{L}(\mathcal{E}^N)} \prod_{k=1}^{N} \mathcal{E}_{\Gamma_{k-1}, \Gamma_k}, \quad (30)$$

where $\sum_{\underline{\Gamma}} \kappa(\underline{\Gamma}) = 1$. Therefore, $\kappa(\underline{\Gamma})$ and $\mathbb{P}(\underline{X})$ can be read respectively as the probability function of $\underline{\Gamma}$ and as a $\kappa(\underline{\Gamma})$-weighted mixture of laws, each defined as the product $\prod_{k=1}^{N} \mathcal{P}_{\Gamma_{k-1}, \Gamma_k}(x_k)$.

Second, let $\Gamma_{\ge t}$ denote the set of chains starting at index t and stopping at index $N - 1$: $\Gamma_{\ge t} \equiv (\Gamma_t, \ldots, \Gamma_{N-1})$. For a
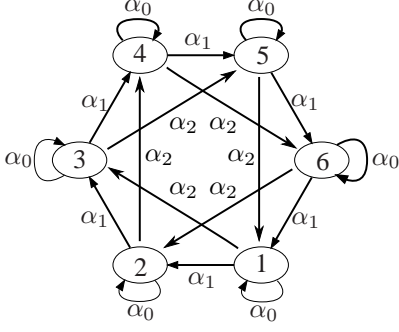
Fig. 1. **Transition graph.** Example for $d = 6$ and $\mathcal{E} = \alpha_0 I_d + \alpha_1 J_d + \alpha_2 J_d^2$ ($\alpha_0 + \alpha_1 + \alpha_2 = 1$).

given pair (Γ_0, Γ_N), Eq. (30) shows that:

$$\mathbb{P}(\Gamma_k = j | \Gamma_0 = \gamma_0, \ldots \Gamma_{k-1} = \gamma_{k-1}) =$$

$$\frac{\mathcal{E}_{\gamma_0,\gamma_1} \ldots \mathcal{E}_{\gamma_{k-1},j} \sum_{\Gamma_{\geq k+1}} \mathcal{E}_{j,\Gamma_{k+1}} \ldots \mathcal{E}_{\Gamma_{N-1},\Gamma_N}}{\mathcal{E}_{\gamma_0,\gamma_1} \ldots \mathcal{E}_{\gamma_{k-2},\gamma_{k-1}} \sum_{\Gamma_{\geq k}} \mathcal{E}_{\gamma_{k-1},\Gamma_k} \ldots \mathcal{E}_{\Gamma_{N-1},\Gamma_N}}$$

$$= \mathcal{E}_{\gamma_{k-1},j} \frac{\left(\mathcal{E}^{N-k}\right)_{j,\Gamma_N}}{\left(\mathcal{E}^{N-k+1}\right)_{\gamma_{k-1},\Gamma_N}} = \mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = \gamma_{k-1})$$
(31)

and hence that $\underline{\Gamma}$ consists of an inhomogeneous d-state Markov chain, with transition probability matrix at step k reading:

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j} \frac{\left(\mathcal{E}^{N-k}\right)_{j,\Gamma_N}}{\left(\mathcal{E}^{N-k+1}\right)_{i,\Gamma_N}}. \quad (32)$$

This shows that $\mathcal{E}$ can now be interpreted as (the basis for) the transition matrix underlying the Markov chain $\underline{\Gamma}$, as illustrated in Fig. 1.

Third, Eq. (30) enables us to show that the initial distribution for (Γ_0, Γ_N) reads:

$$\mathbb{P}(\Gamma_0 = i, \Gamma_N = j) = \frac{\mathcal{A}_{i,j}(\mathcal{E}^N)_{i,j}}{\mathcal{L}\left(\mathcal{E}^N\right)}. \quad (33)$$

Combined together, these three steps enable us to synthesize numerically Vector $\underline{X}_k$, as defined from Eq. (2), using a Hidden Markov Model, with 4 hidden states: the current state Γ_k, the previous state Γ_{k-1} and the pair formed by the initial and final states (Γ_0, Γ_N). Combining Eqs. (32) and (33), the synthesis algorithm can be sketched as follows:

Step1: Initialization:
Use Eq. (33) to generate the states Γ_0 and Γ_N.
Step2: Iteration on $k = 1, \ldots, N$:

Step2.1: Choose at random state Γ_k, according to the transition probability given in Eq. (32);
Step2.2: Generate X_k according to $\mathcal{P}_{\Gamma_{k-1},\Gamma_k}$.

Under the particular choices $\mathcal{A}_{i,j} = 1/d$ (cf. Eq. (14)) and $\mathcal{E}$ doubly stochastic (cf. Eq. (15)), it has been shown that there is no need to set a priori the final state Γ_N and that the Markov chain becomes homogeneous, hence that Eqs. (32) and (33) can be rewritten as (cf. [18]):

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j} \text{ and} \quad (34)$$

$$\mathbb{P}(\Gamma_0 = i) = \frac{1}{d}. \quad (35)$$

Such simplifications show that $\mathcal{E}$ exactly defines the transition matrix for the Markov chain Γ and, moreover, that the initial state of the Markov chain Γ_0 follows a uniform distribution.

The computational cost of the synthesis procedure can be evaluated as follows. For the general case where $\mathcal{E}$ is not an invertible matrix, the algorithmic complexity is dominated by the cost of the computation of matrix powers, and hence a total complexity in $\mathcal{O}(d^3 N \ln N)$ (with a $\mathcal{O}(d^3)$ matrix multiplication cost and a $\mathcal{O}(\ln N)$ matrix power computation cost). When $\mathcal{E}$ is invertible, the global cost reduces to $\mathcal{O}(d^3 N)$ (as matrix multiplications only are required). With the choice of doubly stochastic matrices $\mathcal{E}$ and $\mathcal{A}_{i,j} = 1/d$, the computational cost is further reduced to $\mathcal{O}(N \ln d)$. Therefore, the synthesis procedure scales efficiently for large signal (large N) and large complex dependence structure (large d) (with numerous further potential optimizations for specific cases such as sparse structure matrix).

B. Design

Let us now address the issue of how to select the elements constitutive of the model, given a targeted $\underline{X}_N$.

Prescribing directly a targeted joint distribution function consists of a difficult theoretical problem, that also often does not match application purposes. Instead, it is often natural to split this general question into two simpler sub-problems: on one hand, fixing the dependence (or covariance) structure; on the other hand, fixing the marginal distribution(s). Sections II and III indicate that $\binom{r_M}{2}$ defines the (maximal) number of time-scales involved in the dependence (covariance) structure function, while the joint choice of $\mathcal{A}$ and $\mathcal{E}$ controls the shape of the dependence (e.g., stationarity, cross-covariance, . . .). Examples of construction enabling us to reach targeted dependencies are detailed in Section V.

Let us for now assume that d, $\mathcal{A}$, and $\mathcal{E}$ are chosen (hence that the dependence structure is fixed) and let us concentrate here on designing Matrix $\mathcal{P}$, so as to reach targeted marginal distributions. Also, for the sake of simplicity, let us concentrate on the stationary case first. Extensions to the non-stationary case will then be discussed.

It is first worth emphasizing that prescribing marginals and dependence cannot be treated as fully independent problems (only the global structure of the dependence is not related to the choice of Matrix $\mathcal{P}$). Indeed choosing $\mathcal{P}$ fixes both univariate distributions (according to Eq. (5)) and the coefficients governing the relative amplitude of the dependence structure, through the matrices $M(q)$ (cf. Eq. (9)). Moreover, the prescribed marginal imposes some constraints on the matrix $M(q)$. For instance, Eq. (11) implies that:

$$M(1)_{i,j} < \mathcal{E}_{i,j} \mathbb{E}\left[X | X > x\right], \quad x \equiv F_S^{-1}(1 - c_{i,j}), \quad (36)$$

with X generated according to $\mathbb{P}_S$ (as defined in Eq. (11)), and where F_S denotes the cumulative function associated to the marginal distribution $\mathbb{P}_S$. The difficulty thus consists of devising a constructive procedure for Matrix $\mathcal{P}$ enabling practitioners to reach jointly the targeted marginal distribution

$\mathbb{P}_S$ (according to Eq. (11)) and targeted Matrices $M(q)$ (cf. Eq. (9)). To disentangle both groups of constraints, the $\mathcal{P}_{i,j}$ are parametrized using a matrix of (strictly) positive functions $g_{i,j}(x)$:

$$c_{i,j}\mathcal{P}_{i,j}(x) = \frac{g_{i,j}(x)}{\sum_{l,m} g_{l,m}(x)}\mathbb{P}_S(x). \qquad (37)$$

With this parametrization, Eq. (11) is automatically satisfied. To ensure that the $\mathcal{P}_{i,j}$ are probability distributions, it is needed that

$$\int \frac{g_{i,j}(x)}{\sum_{l,m} g_{l,m}(x)}\mathbb{P}_S(x)dx = c_{i,j}. \qquad (38)$$

Splitting $g_{i,j}$ into $g_{i,j}(x) = \mu_{i,j}h_{i,j}(x)$ permits to use the free parameters $\mu_{i,j}$ to solve Eq. (38), for any fixed $h = (h_{i,j})$:

$$\int \frac{\mu[h]_{i,j}h_{i,j}(x)}{\sum_{l,m}\mu[h]_{l,m}h_{l,m}(x)}\mathbb{P}_S(x)dx = c_{i,j}. \qquad (39)$$

Plugging the solution $(\mu[h]_{i,j})$ of Eq. (39) into Eq. (37) yields:

$$\mathcal{P}[h]_{i,j}(x) = \frac{\mu[h]_{i,j}h_{i,j}(x)}{c_{i,j}\sum_{l,m}\mu[h]_{l,m}h_{l,m}(x)}\mathbb{P}_S(x). \qquad (40)$$

The last step is to find h such that:

$$M(q)_{i,j} = \mathcal{E}_{i,j}\int x^q \mathcal{P}[h]_{i,j}(x)dx \qquad (41)$$

where $M(q)$ are the prescribed moment matrices. Essentially, this parametrization of $\mathcal{P}$ in terms of $h_{i,j}(x)$ enables us to recast Eq. (6) as a non linear equation in $h_{i,j}(x)$ (cf. Eq. (41)), where $h_{i,j}(x)$ are strictly positive functions. To practically find solutions to Eq. (41), the functions $h_{i,j}(x)$ are further constrained to be chosen amongst an arbitrary parametrized kernel family K_p: $h_{i,j}(x) = K_{p_{i,j}}(x)$. A natural choice for the kernel parameters p would be to consider vectorial parameters with a dimension equal to the number of targeted moments orders. For instance, when the targeted $\mathbb{P}_S$ is a normal law and $M(1)$ and $M(2)$ are fixed a priori, it is natural (but not mandatory) to select $K_{m,\sigma} = \mathcal{N}_{m,\sigma}$. Further examples are given in Section V. Within this framework, Eqs. (37) to (41) can be solved numerically step-by-step. Examples of results obtained with the constructive method are detailed in Section V and illustrated in Fig. 4.

In summary, to design Vector $\underline{X}_N$ with prescribed univariate distributions and dependence structure, the following procedure is to be followed:

1) Select d, $\mathcal{E}$ and $\mathcal{A}$ in agreement with the targeted dependence structure.
2) Choose the moment matrices $M(q)$ to determine the dependence coefficients.
3) Choose a kernel family K_p.
4) Solve Eqs. (37) to (41) to compute the distribution matrix $\mathcal{P}$ in agreement with the targeted $\mathbb{P}_k$s.

A MATLAB implementation of this procedure is available upon request. The potential of this method is further explored in Section V where simple (hence pedagogical) examples are devised and discussed.

For the general non-stationary case, the above algorithm cannot be used as is and a general solution is still under construction. However, for the restriction of the non-stationary case to the fairly broad and practically efficient case of block structured matrices, detailed in Section V-B, the above algorithm can straightforwardly be applied independently for each component of the targeted random vector. As illustrated in Section V-B, both the block structured matrix framework and the above algorithm provide practitioners with a rather general tool to define and synthesize numerically a large number of realizations of random vectors.

V. ILLUSTRATIONS

This sections aims at illustrating the potential of the general construction above in two specific contexts: First, we concentrate on stationary vectors $\underline{X}_N$, i.e., on stationary time series synthesis; Second, we illustrate for trivariate random vectors how to vary the joint distribution functions while maintaining fixed the three different marginal distributions.

A. Stationary time series

1) Circular structure matrix: To design stationary time series, elaborating on a preliminary work (cf. [18]), we now select $\mathcal{A}_{i,j} = (1/d)$ and a particularly fruitful choice of doubly stochastic matrices:

$$\mathcal{E} = \sum_{k=0}^{d-1}\alpha_k J_d^k, \quad \sum_{k=0}^{d-1}\alpha_k \equiv 1, \qquad (42)$$

where $J_d \in M_d(\mathbb{R})$ is defined as:

$$J_d = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

and $J_d^0 = I_d$, the Identity matrix. It is easy to verify that $[\mathcal{A}^T, J_d] = 0$. Therefore, $[\mathcal{A}^T, \mathcal{E}] = 0$ and $\underline{X}_N$ is stationary. Moreover, the eigenvalues of J_d are easily computed as functions of the roots of order d of the unity, $\omega = \exp(2\imath\pi/d)$, which motivates the use of this matrix:

$$\lambda_i = \sum_{k=0}^{d-1}\alpha_k\omega^{ik}, \ i = 0,\ldots,d-1. \qquad (43)$$

The eigenvectors of J_d also have simple expressions:

$$B_{i,j} = \omega^{ij}. \qquad (44)$$

From these definitions, it is useful to change of basis:

$$\tilde{\mathcal{A}} = B\mathcal{A}B^{-1} = \begin{pmatrix} 0 & \cdots & & 0 \\ \vdots & & & \cdots \\ 0 & & \vdots & 1 \end{pmatrix}$$

$$\tilde{\mathcal{E}} = B^{-1}\mathcal{E}B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \qquad (45)$$

$$\tilde{M}(q) = B^{-1}M(q)\,B$$

Such calculations lead to a simple expression of the dependencies:

$$\mathbb{E}\left[X_k^{q_1} X_l^{q_2}\right] = \sum_i \lambda_i^{k-l} \tilde{M}(q_1)_{d,i} \tilde{M}(q_2)_{i,d} \qquad (46)$$

For ease of notations, let us define the two vectors $(C_{\tilde{M}(q)})_k = \sum_i \tilde{M}(q)_{i,k}$, $(R_{\tilde{M}(q)})_k = \sum_j \tilde{M}(q)_{k,j}$ and let $\mathcal{F}$ denote the (normalized) discrete Fourier transform :

$$\mathcal{F}(v)_k = \frac{1}{d}\sum_{l=1}^{d} v_l e^{\frac{2i\pi kl}{d}} = \sum_l v_l \omega^{kl}$$

This leads to:

$$\mathbb{E}\left[X_0^q X_t^q\right] - \mathbb{E}\left[X_0^q\right]\mathbb{E}\left[X_t^q\right] =$$
$$\sum_{k=1}^{\lfloor d/2 \rfloor} m_k \mathfrak{Re}\left\{ \mathcal{F}(R_{\tilde{M}(q)})_k \overline{\mathcal{F}(C_{\tilde{M}(q)})_k} e^{-\frac{t-1}{\tau_k}} e^{i\frac{2\pi(t-1)}{T_k}} \right\}, \qquad (47)$$

where $\lfloor z \rfloor$ stands for the integer part of z, $m_k = 1$ if $2k = d$ and $m_k = 2$ otherwise, and $\mathfrak{Re}\left\{\right\}$ denotes the real part. In this specific case, there is thus at most $\lfloor d/2 \rfloor + 1$ distinct time scales in the signal. As a pedagogical example, choosing $\mathcal{E} = \alpha_0 I + \alpha_1 J_d$ (where $\alpha_0 + \alpha_1 = 1$) enables to further study the form of the covariance function. The eigenvalues of $\mathcal{E}$ read $\lambda_k = \alpha_0 + \alpha_1 e^{\frac{2i\pi k}{d}}$ and can be rewritten as $\lambda_k = e^{-\frac{1}{\tau_k}} e^{\pm i\frac{2\pi}{T_k}}$, $k = 1, \ldots \lfloor d/2 \rfloor$. Combining these results with Eq. (47) shows that $\tau_k = -1/\ln|\lambda_k|$ and $T_k = 2\pi/\arg(\lambda_k)$ are characteristic dependence time scales and periods that depend on the joint choice of d and α_0. It can also be shown that $\tau_k \approx_{\alpha_0\alpha_1 \to 0} \left[\alpha_0\alpha_1\left(1 - \cos\frac{2\pi k}{d}\right)\right]^{-1}$. Moreover, increasing d increases both the number of distinct dependence time scales and the ratio of the smallest to the largest such characteristic time scales, which can be shown to vary asymptotically ($\alpha_0\alpha_1 \to 0$) as $(d/2\pi)^2/2$.

2) Example: To illustrate the potential of the proposed time series theoretical construction and synthesis procedure, a pedagogical example is devised. First, it is intended to design a stationary Gaussian time series X of size N, with a Gaussian marginal distribution $P(x) = \mathcal{N}_{0,1}$, an auto-covariance function that consists of a δ function (i.e., no correlation and $M(1) \equiv 0$), but dependence as fixed here by the 4-th order statistics, i.e., by prescribing $M(2)$. By definition, X is hence not a jointly Gaussian process (i.e., its joint distribution departs from a multivariate Gaussian law). Second, it is intended, using the same d, $\mathcal{A}$ and $\mathcal{E}$, to design a second time series Y, independent of X, but with same size N, same marginal distribution and autocovariance function but different dependence (via a different $M(2)$ matrix) and hence a joint distribution different from that of X.

For these constructions, $d = 6$ is selected to obtain 3 distinct time scales and $\mathcal{E} = \alpha_0 I_d + \alpha_1 J_d$ ($\alpha_0 + \alpha_1 = 1$) for the sake of simplicity. With these choices, the three available time scales read:

$$\tau_1 \underset{\alpha_1 \to 0}{\approx} \frac{2}{\alpha_1}, \ \tau_2 \underset{\alpha_1 \to 0}{\approx} \frac{2}{3\alpha_1}, \ \tau_3 \underset{\alpha_1 \to 0}{\approx} \frac{1}{2\alpha_1}. \qquad (48)$$

The two time series X and Y differ only in the autocovariance of their squares ($\mathbb{E}\left[X_0^2 X_t^2\right] - \mathbb{E}\left[X_0^2\right]\mathbb{E}\left[X_t^2\right]$). This implies that the matrices $M(2)$ of the two signals differ. An
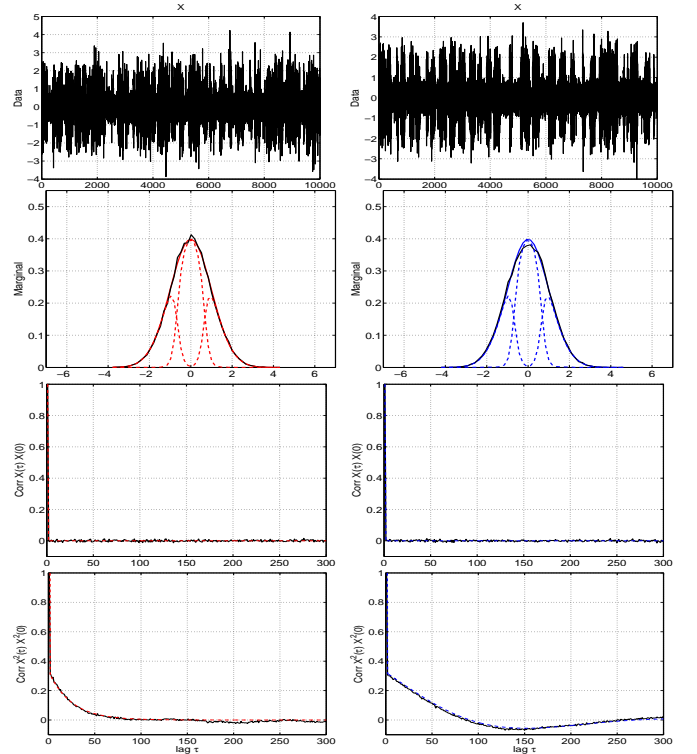

Fig. 2. **Numerical Synthesis.** Two different time series with the same marginal distribution (mixture of two Gaussians, $\sigma_1 = 0.5$, $\sigma_2 = 2$), the same covariance functions (chosen as δ functions) but different covariance functions for their squares, hence different prescribed joint distributions ($\alpha_0 = 0.98$). Left side $\underline{X}$, right side $\underline{Y}$. First line, one realization of the time series. Then, from top to bottom, estimated (solid black lines) and theoretical (dashed colored lines) marginals, correlation functions and correlation functions for the squared time series.

interesting simplification at this step is to choose the matrices $M(2)$ to be of the form $M(2) =, \alpha_0 D + \alpha_1 D J_d$ where D is diagonal. In this case Eq. (47) further simplifies to:

$$\mathbb{E}\left[X_0^q X_t^q\right] - \mathbb{E}\left[X_0^q\right]\mathbb{E}\left[X_t^q\right] =$$
$$\sum_{k=1}^{\lfloor d/2 \rfloor} m_k |\mathcal{F}(\mathrm{diag}(D))_k|^2 \mathfrak{Re}\left\{(\alpha_0 + \alpha_1\omega^k)e^{-\frac{t-1}{\tau_k}} e^{2\pi i \frac{t-1}{T_k}} \right\}, \qquad (49)$$

where $diag(D)_i = D_{i,i}$ is the diagonal vector of D. Therefore, choosing for X,

$$D_{i,i} = \begin{cases} \sigma_1 & \text{i even} \\ \sigma_2 & \text{otherwise} \end{cases}, \quad \sigma_1 + \sigma_2 = 1 \qquad (50)$$

induces that τ_3 only, i.e., the shortest time scale, appears in the auto-covariance of X^2:

$$\mathbb{E}\left[X_0^2 X_t^2\right] - \mathbb{E}\left[X_0^2\right]\mathbb{E}\left[X_t^2\right] = \frac{(\sigma_1 - \sigma_2)^2}{4}(\alpha_0 - \alpha_1)^t. \qquad (51)$$

Conversely, imposing for Y

$$D_{i,i} = \begin{cases} \sigma_1 & i \le 3 \\ \sigma_2 & i > 3 \end{cases}, \quad \sigma_1 + \sigma_2 = 1 \qquad (52)$$

leads to the fact that both τ_1 and τ_3 contribute to the autocovariance of Y^2:

$$\mathbb{E}\left[Y_0^2 Y_t^2\right] - \mathbb{E}\left[Y_0^2\right]\mathbb{E}\left[Y_t^2\right] =$$

$$\frac{(\sigma_1 - \sigma_2)^2}{36}\left[(\alpha_0 - \alpha_1)^t + 4\mathfrak{Re}\left\{(2 - \alpha_1 - \sqrt{3}\iota\alpha_1)\lambda_1^{t-1}\right\}\right].$$
(53)

However, $\tau_1 \approx 4\tau_3$ is dominant at large t and hence constitutes the leading term.

With these pedagogical and documented choices of $M(2)$ for X and Y, we can use the numerical procedure devised in Section IV-B, with a Gaussian kernel $K_{m,\sigma}(x) = \exp(-(x-m)^2/(2\sigma^2))$, to compute the associated distributions $\mathcal{P}_{i,j}$. The hidden Markov chain procedure described in Section IV-A is then used to synthesize the signals X and Y. The analysis of these synthesized signals produces the results reported in Fig. 2. This figure presents for X (left column) and Y (right column) a particular realization of the time series, the estimated and targeted univariate distributions, covariance functions and covariance functions for the squared time series (from top to bottom). It clearly shows that X and Y have the same marginal and covariance but different joint distributions (as targeted). Though sharing the same pedagogical goal, the examples devised here significantly differ from those presented in [18] as both the targeted marginal and the constructive Kernel differ.

Using the same construction procedure, other examples with the same (non necessarily Gaussian) marginals, same (non necessarily δ-) autocovariance functions but different joint distributions could as easily be devised and are available upon request.

B. Random vector

1) Multivariate design: Let us now consider the design of a random vector whose components have different univariate distributions and are dependent. As seen in Section III, for a fixed d, when N increases, $\underline{X}_N$ tends to become stationary. A simple way around this drawback is to increase the size of R_d with the size of $\underline{X}_N$, by choosing $d = (N+1)d^*$, while keeping a block triangular superior structure, to avoid a too large increase in the number of degrees of freedom:

$$R_d(x) = \begin{pmatrix} 0_{d^*} & R_{d^*}^{(1)}(x) & & 0_{d^*} \\ & 0_{d^*} & \ddots & \\ & & 0_{d^*} & R_{d^*}^{(N)}(x) \\ 0_{d^*} & & & 0_{d^*} \end{pmatrix}, \quad (54)$$

where 0_{d^*} is a 0-block of size $d^* \times d^*$ and $R_{d^*}^{(k)}$ are $d^* \times d^*$ matrices, as defined in Eq. (2):

$$R_{d^*}^{(k)}(x) = \mathcal{E}^{(*k)}\mathcal{P}^{(*k)}(x), \quad \int_{-\infty}^{+\infty}\mathcal{P}_{i,j}^{(*k)}(x)dx = 1.$$

Let us also choose the projection matrix $\mathcal{A}$ such that

$$\mathcal{A} = \begin{pmatrix} 0_{d^*} & \dots & 0_{d^*} \\ \vdots & & \vdots \\ \mathcal{A}^* & \dots & 0_{d^*} \end{pmatrix}, \quad (55)$$

where $\mathcal{A}^*$ is a positive entry matrix of size $d^* \times d^*$. If one further defines $\mathcal{L}^*(M) = \mathrm{tr}\left(\mathcal{A}^{*T}M\right)$ (as a linear form on $d^* \times d^*$ matrices), Eq. (3) becomes:

$$\mathbb{P}(x_1, \dots, x_N) = \frac{\mathcal{L}^*\left(\prod_{k=1}^{N} R_{d^*}^{(k)}(x_k)\right)}{\mathcal{L}^*\left(\prod_{k=1}^{N}\mathcal{E}^{(*k)}\right)}.$$

By further restricting to the case $\mathcal{E}^{(*k)} = \mathcal{E}^*$, the joint distribution simplifies to:

$$\mathbb{P}(x_1, \dots, x_N) = \frac{\mathcal{L}^*\left(\prod_{k=1}^{N} R_{d^*}^{(k)}(x_k)\right)}{\mathcal{L}^*\left(\mathcal{E}^{*N}\right)}. \quad (56)$$

The joint distribution in Eq. (56) consists of a variation on Eq. (3), where the constant probability matrix $\mathcal{P}$ has been replaced by a varying probability matrix $\mathcal{P}^{(*k)}$. The general formulation in Eq. (54) and Eq. (55) shows that this particular setting is nothing but a convenient notation that however corresponds to a subcase of the general framework: Therefore all results developed in Sections II to IV remain valid.

It is hence straightforward to derive the corresponding expressions for the univariate distributions, the p-samples probability distributions and moments:

$$\mathbb{P}_k(X_k = x_k) = \frac{1}{\mathcal{L}\left(\mathcal{E}^{*N}\right)}\mathcal{L}\left(\mathcal{E}^{*k_1 - 1}R_{d^*}^{(k)}(x_k)\mathcal{E}^{*N-k}\right)$$
(57)

$$\mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) = \frac{1}{\mathcal{L}\left(\mathcal{E}^{*N}\right)}$$
$$\mathcal{L}\left(\mathcal{E}^{*k_1 - 1}\left(\prod_{r=1}^{p-1}R_{d^*}^{(k_r)}(x_{k_r})\mathcal{E}^{*k_{r+1}-k_r-1}\right)R_{d^*}^{(k_p)}(x_{k_p})\mathcal{E}^{*N-k_p}\right)$$
(58)

$$\mathbb{E}\left[\prod_{r=1}^{p}X_{k_r}^{q_r}\right] = \frac{1}{\mathcal{L}(\mathcal{E}^{*N})}$$
$$\mathcal{L}\left(\mathcal{E}^{*k_1 - 1}\left(\prod_{r=1}^{p-1}M^{(*k_r)}(q_r)\mathcal{E}^{*k_{r+1}-k_r-1}\right)M^{(*k_p)}(q_p)\mathcal{E}^{*N-k_p}\right),$$
(59)

with $M^{(*k)}(q) = \int_{-\infty}^{+\infty}x^q R_{d^*}^{(k)}(x)dx$. Using this particular structure for the matrix R_d permits to define each component of the random vector $\underline{X}_N$ with a relative independence. Notably, the univariate distribution of X_k depends only on $\mathcal{P}^{(*k)}$. Moreover, with these matrices $\mathcal{A}$ and R_d, the hidden Markov chain Γ starts in the first upper diagonal block of size d^* and must end in the last upper diagonal block. In other words, at each step, the hidden Markov chain goes from the kth block to the $(k+1)^{th}$ block using the transition i, j which corresponds to a probability law belonging to $R_{d^*}^{(k)}$.

2) Trivariate Example: Let us now develop the construction of two different trivariate random vectors $\underline{X}_N = (X_1, X_2, X_3)$ and $\underline{Y}_N = (Y_1, Y_2, Y_3)$, with marginal distributions set to a Gaussian distribution $\mathcal{N}_{0,1}$, for X_1 and Y_1, to a Gamma distribution, with shape parameter $\alpha = 2$ and scale parameter $\beta = 1$ for X_2 and Y_2, and to a Gamma distribution with
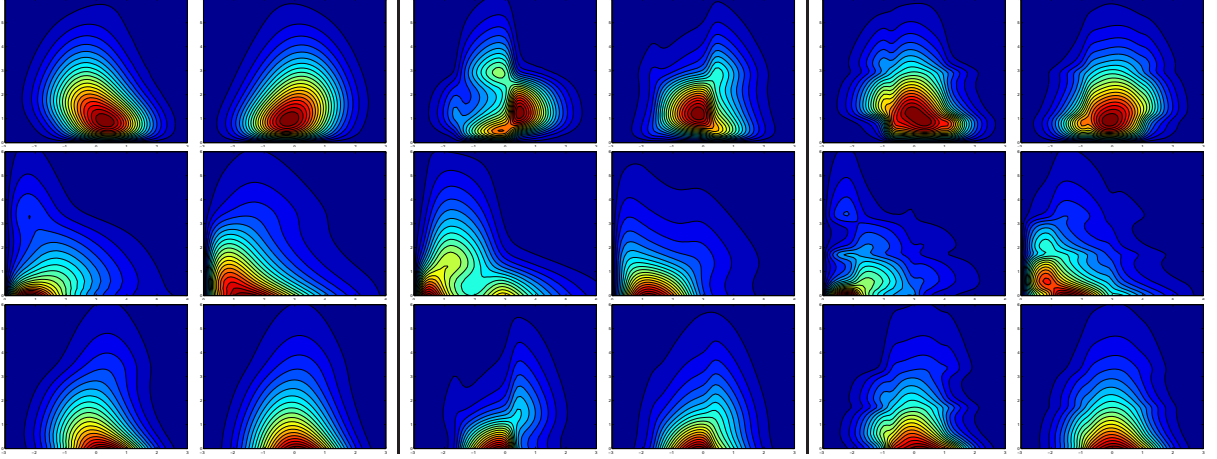
Fig. 3. Bivariate probability distributions of $\underline{X}_N$ (Left) and $\underline{Y}_N$ (Middle) and $\underline{Z}_N$(Right), for Pairs 1-2 (top), Pairs 2-3 (middle), Pairs 1-3 (bottom). For each vector: Left: $p = 0.1$. Right: $p = 0.8$.
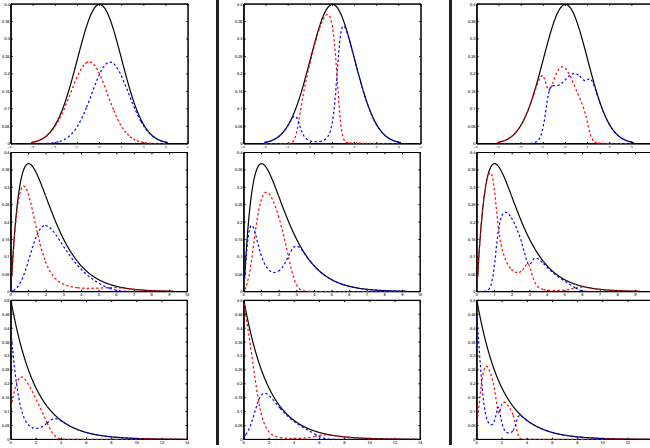

Fig. 4. Marginal distributions $\mathbb{P}_k$ (solid lines) for $\underline{X}_N$ (left), $\underline{Y}_N$ (middle) and $\underline{Z}_N$ (right) and the designed $c_{i,j,k}\mathcal{P}_{i,j}$ (dashed lines); for Components X_1, Y_1, Z_1 (top); Components X_2, Y_2, Z_2 (middle); Component X_3, Y_3, Z_3 (bottom).

$\alpha = 1$ and $\beta = 2$ for X_3 and Y_3. To illustrate the potential of the tool, $\underline{X}_N$ and $\underline{Y}_N$ also have the same correlations, but different joint distributions.

First, $d^* = 2$, $\mathcal{E}^* = \alpha_0 I_d + \alpha_1 J_d$ and $\mathcal{A}^*_{i,j} = (1/d^*)$ are selected. Second, to control correlations, the moment matrix $M^{(*k)}(1)$ is set, for both vectors, to ($k = 1, 2, 3$):

$$M^{(*k)}(1) = \mathcal{E}^* \otimes \begin{pmatrix} m_{k,1} & m_{k,1} \\ m_{k,2} & m_{k,2} \end{pmatrix}, \qquad (60)$$

with the constraints:

$$m_{k,1} + m_{k,2} = 2\mathbb{E}[X_k].$$

Defining $\Delta_k = m_{k,1} - m_{k,2}$, the covariance reads

$$\text{Cov}[X_1, X_2] = \text{Cov}[Y_1, Y_2] = (1 - 2\alpha_0)\Delta_1\Delta_2,$$

$$\text{Cov}[X_2, X_3] = \text{Cov}[Y_2, Y_3] = (1 - 2\alpha_0)\Delta_2\Delta_3,$$

$$\text{Cov}[X_1, X_3] = \text{Cov}[Y_1, Y_3] = (1 - 2\alpha_0)^2\Delta_1\Delta_3.$$

Therefore, the covariance for any two consecutive components depends linearly on α_0, and, when $\Delta_k\Delta_l > 0$ is maximum for $\alpha_0 = 1$, vanishes at $\alpha_0 = 0.5$ and is minimum for $\alpha_0 = 0$.

The two trivariate joint distributions can now be made different via their moment matrices, $M^{(*k)}(q)$, of order $q = 2$, which for $\underline{X}_N$ is set to:

$$M^{(*1)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \, M^{(*3)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 4.5 & 4.5 \\ 11.5 & 11.5 \end{pmatrix},$$

$$M^{(*2)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 4.5 & 4.5 \\ 7.5 & 7.5 \end{pmatrix}; \qquad (61)$$

and for $\underline{Y}_N$ to:

$$M^{(*1)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 0.5 & 0.5 \\ 1.5 & 1.5 \end{pmatrix}, \quad M^{(*3)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix}.$$

$$M^{(*2)}(2) = \mathcal{E}^* \otimes \begin{pmatrix} 2.75 & 2.75 \\ 9.25 & 9.25 \end{pmatrix}. \qquad (62)$$

To construct the $\mathcal{P}^{(*k)}_{i,j}$ from the procedure developed in Section IV-B, a Gaussian kernel $K_{m,\sigma}(x) = \exp(-(x - m)^2/(2\sigma^2))$ is chosen. The resulting probability matrices $\mathcal{P}^{(*k)}$ have only 2 distinct components. Fig. 4 illustrates how the weighted density $c_{i,j,k}$ are combined in order to obtain the marginal distribution $\mathbb{P}_k$ for $\underline{X}_N$ (left) and $\underline{Y}_N$ (middle).

Furthermore, changing the Kernel K parametrizing $\mathcal{P}^{(*k)}_{i,j}$, also constitutes an efficient way to further vary the joint distributions, hence introducing further versatility in the procedure. For example, K could be chosen as the Gamma distribution family, or the union of the Gaussian and Gamma families. To illustrate this, let us now construct a third trivariate random vector $\underline{Z}_N$, sharing the same marginals and covariance $\underline{X}_N$ and $\underline{Y}_N$ (it actually shares exactly the same matrices $M^{(*k)}(1)$ and $M^{(*k)}(2)$ as those of $\underline{X}_N$), though obtained from Kernel $K_{m,\sigma}(x) = (0.1 + ((x - m)/\sigma)^2) \exp(-((x - m)/\sigma)^2)$.

Bivariate partials distributions (rather than trivariate joint distributions, for clarity) are shown in Fig. 3 for the pairs (X_1, X_2) (top row), (X_2, X_3) (middle row) and (X_1, X_3) (bottom row) for $\underline{X}_N$ (left block), $\underline{Y}_N$ (middle block) and

$\underline{Z}_N$ (right block), for negative (left column in each block) and positive (right column) correlations. These plots clearly show that, for any two pairs, the bivariate distributions (hence a fortiori the joint trivariate distributions) are different for the three vectors, though their components have same univariate marginals and same correlations. Furthermore, Fig. 4 shows how the targeted marginal distributions $\mathbb{P}_k$ are obtained by summation of the $c_{i,j,k}\mathcal{P}_{i,j}^{(*k)}$ which were produced by the algorithmic procedure described in Section IV-B. Note that with the chosen setting of the proposed examples, the summation has been a priori limited (for simplicity) to only 2 different $\mathcal{P}_{i,j}^{(*k)}$, while $d^* \times d^* = 4$ distinct distributions would have been available in general. Interestingly, comparisons for a given component (i.e., for a given row) amongst the three vectors illustrates that a same marginal distribution $\mathbb{P}_k$ is attained from different summation of the $c_{i,j,k}\mathcal{P}_{i,j}^{(*k)}$ for the three vectors, which constitutes a signature of the fact that the joint distributions of $\underline{X}_N$, $\underline{Y}_N$ and $\underline{Z}_N$ are different.

Here, the example was chosen trivariate, as a trade-off between ease of exposition (3 partial distributions of pairs of components remain easy to display) and demonstration of the potential and richness of the tool. Multivariate examples are however just as easy to construct.

VI. CONCLUSION AND PERSPECTIVES

A general framework for the theoretical definition and the practical numerical synthesis of random vectors and random time series, with a priori prescribed statistical properties, has been fully worked out, based on a matrix product formalism, and inspired from out-of-equilibrium statistical physics models. Its ability to shape jointly marginal distributions and the dependence structure has been studied both theoretically and practically. Pedagogical examples illustrated the versatility and richness of the procedure in actually attaining targeted properties. Remapping this matrix product framework onto that of Hidden Markov Models enabled us to devise an efficient practical numerical synthesis algorithm. Also, a design procedure enabling to tune the elements of the models so as to reach desired targets has been obtained. Both for design and numerical synthesis, MATLAB implementation of the described procedures are available upon request.

Comparisons of the proposed approach to other numerical synthesis frameworks in terms of potential, versatility, efficiency, precision and implementation are under current investigations but are beyond the scope of the present contribution. Further, the benefits of using other matrices $\mathcal{A}$ and $\mathcal{E}$ will be explored. Moreover, having maintained the writing of the multivariate distributions as a product, as is the case for independent components, leads to possible computations of the distribution of the maximum $W = \max X_i$ or sum $S = \sum X_i$, of the components of $\underline{X}_N$. Such results are of premier importance for the use of such models in statistical physics applications as well as in signal processing for problems involving statistical properties of extremes or time-averages as ensemble average estimators. This is being investigated. To finish with, the potential use of this synthesis tool to generate independent copies of sets of hyper-parameters

in Monte Carlo Markov Chain numerical schemes constitutes a natural track to investigate.

REFERENCES

[1] B. A. Berg, *Markov chain Monte Carlo simulations and their statistical analysis*, World Scientific, 2004.

[2] D.G. Tzikas, A.C. Likas, and N.P. Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[4] C.R. Dietrich and G.N. Newsam, "Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix," *SIAM J. Sci. Comput.*, vol. 18, no. 4, pp. 1088–1107, 1997.

[5] A. Dembo, C.L. Mallows, and L.A. Shepp, "Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with applications to covariance estimation," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1206–1212, 1989.

[6] R.B. Davies and D.S. Harte, "Tests for Hurst effect," *Biometrika*, vol. 74, no. 4, pp. 95–101, 1987.

[7] A.T.A Wood and G. Chan, "Simulation of stationary Gaussian processes in $[0, 1]^d$," *J. of Comput. and Graph. Stat.*, vol. 3, no. 4, pp. 409–432, 1994.

[8] G. Chan and A.T.A. Wood, "Simulation of stationary Gaussian vector fields," *Statistics and Computing*, vol. 9, no. 4, pp. 265–268, 1999.

[9] H. Helgason, V. Pipiras, and P. Abry, "Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding," *Signal Processing*, vol. 91, no. 5, pp. 1123–1133, 2011.

[10] M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice-Hall, 1995.

[11] H. Helgason, V. Pipiras, and P. Abry, "Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant embedding," *Signal Processing*, vol. 91, no. 8, pp. 1741–1758, 2011.

[12] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3-4, pp. 346–382, 2000.

[13] P. Borgnat, P. Abry, and P. Flandrin, "Using surrogates and optimal transport for synthesis of stationary multivariate series with prescribed covariance function and non-gaussian joint-distribution," in *Proceeding of the IEEE Int. Conf. on Acoust. Speech and Sig. Proc. (ICASSP)*, Kyoto, 2012.

[14] M.-H. Chen, Q.-M. Shao, and J. G. Ibrahim, *Monte Carlo methods in Bayesian computation. Springer Series in Statistics*, Springer-Verlag, New York, 2000.

[15] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, "Exact solution of a 1d asymmetric exclusion model using a matrix formulation," *J Phys. A*, vol. 26, pp. 1493–1517, 1993.

[16] K. Mallick and S. Sandow, "Finite dimensional representations of the quadratic algebra: Applications to the exclusion process," *J. Phys. A: Math. Gen.*, vol. 30, pp. 4513, 1997.

[17] R A Blythe and M R Evans, "Nonequilibrium steady states of matrix-product form: a solver's guide," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 46, pp. R333–R441, 2007.

[18] F. Angeletti, E. Bertin, and P. Abry, "Matrix products for the synthesis of stationary time series with a priori prescribed joint distributions," in *Proceeding of the IEEE Int. Conf. on Acoust. Speech and Sig. Proc. (ICASSP)*, 2012.